

# Random Sampling of Sparse Trigonometric Polynomials II - Orthogonal Matching Pursuit versus Basis Pursuit

Stefan Kunis, Holger Rauhut

April 19, 2006; revised February 2, 2008

## Abstract

We investigate the problem of reconstructing sparse multivariate trigonometric polynomials from few randomly taken samples by Basis Pursuit and greedy algorithms such as Orthogonal Matching Pursuit (OMP) and Thresholding. While recovery by Basis Pursuit has recently been studied by several authors, we provide theoretical results on the success probability of reconstruction via Thresholding and OMP for both a continuous and a discrete probability model for the sampling points. We present numerical experiments, which indicate that usually Basis Pursuit is significantly slower than greedy algorithms, while the recovery rates are very similar.

**Key Words:** random sampling, trigonometric polynomials, Orthogonal Matching Pursuit, Basis Pursuit, Thresholding, sparse recovery, random matrices, fast Fourier transform, nonequispaced fast Fourier transform

**AMS Subject classification:** 94A20, 42A05, 15A52, 90C05, 90C25

## 1 Introduction

In the last two years the rapidly growing field of compressed sensing has attracted much attention [1, 8, 9, 12, 13, 14, 18, 33, 38]. Its basic idea is that sparse or compressible signals can be reconstructed from vastly incomplete non-adaptive information. By “sparse” we mean that a vector has only few non-zero coefficients, while “compressible” expresses that a vector can be well-approximated by a sparse one.

Previous work on this topic includes the reconstruction of Fourier coefficients from samples taken randomly on a lattice by the *Basis Pursuit* (BP) principle [8, 12]. This consists in minimizing the  $\ell^1$ -norm of the Fourier coefficients subject to the condition that the corresponding trigonometric polynomial matches the sampling points. Indeed, it was proven by Candès, Romberg and Tao in [8] in the setting of the discrete Fourier transform that this scheme recovers the coefficients exactly with high probability provided the number of samples is high enough compared to the sparsity, i.e., the number of non-zero coefficients. This result has been generalized by the second author of the present paper in [33] for the case of samples taken uniformly at random from the cube  $[0, 2\pi]^d$ .

Another line of research suggests greedy methods such as (Orthogonal) Matching Pursuit (OMP) and Thresholding for sparse reconstruction tasks [28, 18, 41, 42]. OMP and Thresholding are conceptually simple to implement and potentially faster than BP. In particular, they

may easily take into account fast algorithms for multiplication with the involved matrices, while most standard software for convex optimization [6, 29] does not allow this.

This paper is devoted to the theoretical and numerical investigation and comparison of Thresholding, OMP and BP for the recovery of sparse trigonometric polynomials from randomly taken samples. Our theoretical results indicate that indeed all the methods are suitable for this task. The novelty in the present paper is a performance analysis for OMP and Thresholding, although the theoretical achievements for OMP are only partial so far. In contrast to BP, the greedy algorithms give only a non-uniform guarantee of recovery at a sufficiently small ratio of the number of samples to the sparsity. This means that they guarantee recovery with high probability only for the given trigonometric polynomial, while BP can actually guarantee recovery of *all* sufficiently sparse trigonometric polynomials from a single sampling set.

In practice however, a non-uniform guarantee might be sufficient. Indeed, our numerical experiments suggest that OMP even slightly outperforms BP on generic signals with respect to reconstruction rate. Considering that greedy algorithms are usually significantly faster than BP one would probably use OMP for most applications despite its lack of uniformity.

For related work on this topic, also known as compressed sensing, we refer to [1, 8, 12, 9, 10, 11, 14, 15, 16, 17, 18, 38] and the references therein. For more information on sampling of (not necessarily sparse) trigonometric polynomials in a probabilistic setting the reader may consult [2, 4, 24].

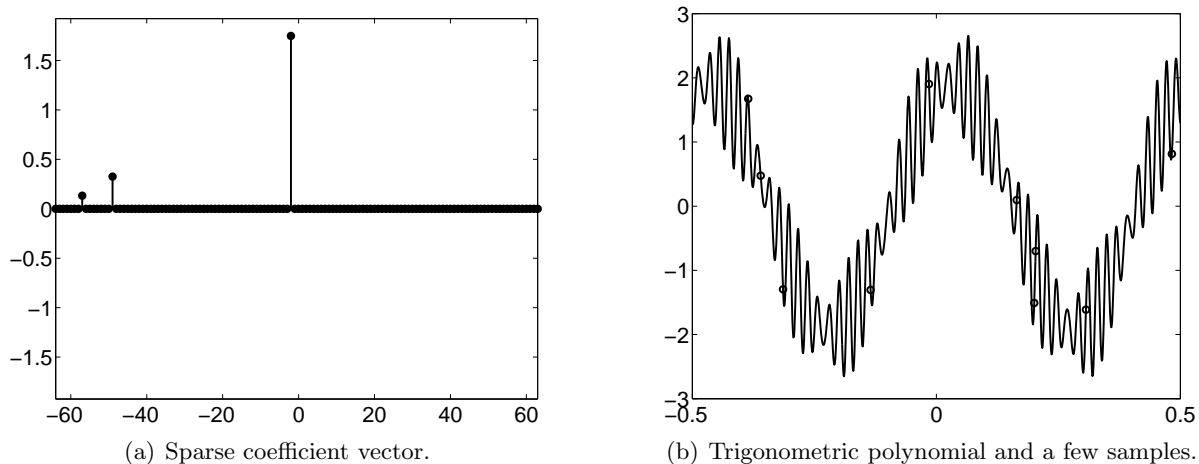


Figure 1: Sparse vector of Fourier coefficients and the corresponding trigonometric polynomial (real part). After sampling at a few randomly chosen points, Orthogonal Matching Pursuit (OMP), i.e., Algorithm 1, as well as the Basis Pursuit principle recover the coefficient vector perfectly with high probability.

The paper is organized as follows: After introducing the necessary notation, including the Orthogonal Matching Pursuit algorithm, we first review known results for Basis Pursuit. Then we present our main result concerning Thresholding and OMP. Based on the coherence parameter we also provide uniform reconstruction results for Thresholding and OMP, cf. Subsection 2.4.

In Section 3 all proofs of the obtained results are given. Section 4 presents extensive numerical experiments. Finally, Section 5 makes conclusions and discusses possible future work.

## 2 Main Results

### 2.1 The Setting

For some finite subset  $\Gamma \subset \mathbb{Z}^d$ ,  $d \in \mathbb{N}$ , we let  $\Pi_\Gamma$  denote the space of all trigonometric polynomials in dimension  $d$  whose coefficients are supported on  $\Gamma$ . Clearly, an element  $f$  of  $\Pi_\Gamma$  is of the form

$$f(x) = \sum_{k \in \Gamma} c_k e^{ik \cdot x}, \quad x \in [0, 2\pi]^d,$$

with some Fourier coefficients  $c_k \in \mathbb{C}$ . The dimension of  $\Pi_\Gamma$  will be denoted by  $D := |\Gamma|$ . Taking  $\Gamma = \{-q, -q+1, \dots, q\}^d$  yields the space  $\Pi_\Gamma = \Pi_q^d$  of all trigonometric polynomials of maximal order  $q$ .

We will mainly deal with “sparse” trigonometric polynomials, i.e., we assume that the sequence of coefficients  $c_k$  is supported only on a set  $T$ , which is much smaller than  $\Gamma$ . However, a priori nothing is known about  $T$  except for its maximum size. Thus, it is useful to introduce the set  $\Pi_\Gamma(M) \subset \Pi_\Gamma$  of all trigonometric polynomials whose Fourier coefficients are supported on a set  $T \subset \Gamma$  satisfying  $|T| \leq M$ . Note that

$$\Pi_\Gamma(M) = \bigcup_{T \subset \Gamma, |T| \leq M} \Pi_T$$

is not a linear space.

Our aim is to sample a trigonometric polynomial  $f$  of  $\Pi_\Gamma(M)$  at  $N$  points  $x_1, \dots, x_N \in [0, 2\pi]^d$  and try to reconstruct  $f$  from these samples. If for some  $m \in \mathbb{N}$  the sampling points are located on the grid

$$\frac{2\pi}{m} \mathbb{Z}_m^d = \left\{ 0, \frac{2\pi}{m}, \dots, \frac{2\pi(m-1)}{m} \right\}$$

then this problem can also be interpreted as reconstructing a sparse vector from partial information on its discrete Fourier transform.

Basis Pursuit consists in solving the following  $\ell^1$ -minimization problem

$$\min \|d\|_1 := \sum_{k \in \Gamma} |d_k| \quad \text{subject to} \quad \sum_{k \in \Gamma} d_k e^{2\pi i k \cdot x_j} = f(x_j), \quad j = 1, \dots, N. \quad (2.1)$$

This task can be performed with convex optimization techniques [6]. For real-valued coefficients (2.1) can be reformulated as a linear program while for complex-valued coefficients we obtain a second order cone program. For both kind of problems standard software exists, such as MOSEK [29] or CVX [21] (internally using SeDuMi [40]) and since recently also L1MAGIC [36] (only for real-valued coefficients).

As an alternative to BP we also use greedy algorithms to recover the Fourier coefficients of  $f$  from few samples. In particular, we study Orthogonal Matching Pursuit (Algorithm 1) as well as the very simple Thresholding algorithm (Algorithm 2). We need to introduce some notation. Let  $X = \{x_1, \dots, x_N\}$  be the set of (random) sampling points. We denote by  $\mathcal{F}_X$  the  $N \times D$  matrix (recall that  $D = |\Gamma|$ ) with entries

$$(\mathcal{F}_X)_{j,k} = e^{ik \cdot x_j}, \quad 1 \leq j \leq N, k \in \Gamma. \quad (2.2)$$

---

**Algorithm 1** OMP

---

Input: sampling set  $X \subset [0, 2\pi]^d$ , sampling vector  $\mathbf{f} := (f(x_j))_{j=1}^N$ , set  $\Gamma \subset \mathbb{Z}^d$ .

Optional: maximum allowed sparsity  $M$  or residual tolerance  $\varepsilon$ .

- 1: Set  $s = 0$ , the residual vector  $r_0 = \mathbf{f}$ , and the index set  $T_0 = \emptyset$ .
- 2: **repeat**
- 3:   Set  $s = s + 1$ .
- 4:   Find  $k_s = \arg \max_{k \in \Gamma} |\langle r_{s-1}, \phi_k \rangle|$  and augment  $T_s = T_{s-1} \cup \{k_s\}$ .
- 5:   Project onto  $\text{span}\{\phi_k, k \in T_s\}$  by solving the least squares problem

$$\|\mathcal{F}_{T_s X} d_s - \mathbf{f}\|_2 \xrightarrow{d_s} \min.$$

- 6:   Compute the new residual  $r_s = \mathbf{f} - \mathcal{F}_{T_s X} d_s$ .
- 7: **until**  $s = M$  or  $\|r_s\| \leq \varepsilon$
- 8: Set  $T = T_s$ , the non-zeros of the vector  $c$  are given by  $(c_k)_{k \in T} = d_s$ .

Output: vector of coefficients  $c$  and its support  $T$ .

---

Then clearly,  $f(x_j) = (\mathcal{F}_X c)_j$  if  $c$  is the vector of Fourier coefficients of  $f$ . Let  $\phi_k$  denote the  $k$ -th column of  $\mathcal{F}_X$ , i.e.,

$$\phi_k = \begin{pmatrix} e^{ik \cdot x_1} \\ \vdots \\ e^{ik \cdot x_N} \end{pmatrix},$$

so  $\mathcal{F}_X = (\phi_{k_1} | \phi_{k_2} | \dots | \phi_{k_D})$ . By

$$(\mathcal{F}_{TX})_{j,k} = e^{ik \cdot x_j}, \quad 1 \leq j \leq N, k \in T. \quad (2.3)$$

we denote the restriction of  $\mathcal{F}_X$  to sequences supported only on  $T$ . Furthermore, let  $\langle \cdot, \cdot \rangle$  be the usual Euclidean scalar product and  $\|\cdot\|_2$  the associated norm. We have  $\|\phi_k\|_2 = \sqrt{N}$  for all  $k \in \Gamma$ , i.e., all the columns of  $\mathcal{F}_X$  have the same  $\ell^2$ -norm. We postpone a detailed discussion on the implementation of Algorithm 1 to Section 4.

Of course, the hope is that running OMP or BP on samples of some  $f \in \Pi_\Gamma(M)$  will recover its Fourier coefficients. To analyze the performance of the algorithms we will use two probabilistic models for the sampling points (one for the continuous case and the other one for the discrete Fourier transform case). This random modeling can be understood in the sense that the sampling set  $X$  is 'generic': reconstruction is allowed to fail for certain choices of  $X$ , as long as the probability of encountering such a pathological case is very small.

We will work with the following two probability models for the sampling points:

- (1) We assume that the sampling points  $x_1, \dots, x_N$  are independent random variables having the uniform distribution on  $[0, 2\pi]^d$ . Obviously, the cardinality of the sampling set  $X = \{x_1, \dots, x_N\}$  equals the number of samples  $N$  with probability 1.
- (2) We suppose that the sampling points  $x_1, \dots, x_N$  have the uniform distribution on the finite set  $\frac{2\pi}{m} \mathbb{Z}_m^d$  for some  $m \in \mathbb{N} \setminus \{1\}$ . It will then always be assumed implicitly that

---

**Algorithm 2** Thresholding

---

Input: sampling set  $X \subset [0, 2\pi]^d$ , sampling vector  $\mathbf{f} := (f(x_j))_{j=1}^N$ , set  $\Gamma \subset \mathbb{Z}^d$ ,  
maximum allowed sparsity  $M$

- 1: Find the indices  $T \subset \Gamma$  corresponding to the  $M$  largest inner products  $\{|\langle \mathbf{f}, \phi_k \rangle|\}_{k \in \Gamma}$ .
- 2: Project onto  $\text{span}\{\phi_k, k \in T_s\}$  by solving the least squares problem

$$\|\mathcal{F}_{TX}d - \mathbf{f}\|_2 \xrightarrow{d} \min.$$

- 3: The non-zero entries of the vector  $c$  are given by  $(c_k)_{k \in T} = d$ .

Output: vector of coefficients  $c$  and its support  $T$

---

$\Gamma \subset \mathbb{Z}_m^d$ . Clearly, our second model aims at studying the problem of reconstructing sparse vectors from (partial) information on its discrete Fourier transform.

Observe that it happens with non-zero probability that some of the sampling points coincide, so the cardinality of the sampling set  $X = \{x_1, \dots, x_N\}$  might be smaller than  $N$ . However, for small  $N \ll D$  this effect will occur rather rarely, and also does not harm our theoretical analysis.

We will often refer to the first model as the “continuous model” while the second will be called the “discrete model”. It turns out that one can treat both probability models in parallel.

## 2.2 Previous results for Basis Pursuit

Based on ideas due to Candès, Romberg and Tao in [8], Rauhut has shown the following result concerning recovery of sparse trigonometric polynomials in [33, Theorem 2.1 and Section 3.6(b)].

**Theorem 2.1.** *Let  $\Gamma \subset \mathbb{Z}^d$  finite and  $T \subset \Gamma$ . Set  $M = |T|$  and  $D = |\Gamma|$ . Let  $X = (x_1, \dots, x_N)$  be chosen according to one of our two probability models. If for some  $\epsilon > 0$  it holds*

$$N \geq CM \log(D/\epsilon)$$

*then with probability at least  $1 - \epsilon$  every trigonometric polynomial  $f \in \Pi_\Gamma(M)$  with Fourier coefficients supported on  $T$  can be recovered from its sample values  $f(x_j), j = 1, \dots, N$ , via Basis Pursuit. The constant  $C$  is absolute.*

In particular, if the sparsity  $M$  is small and the dimension  $D$  large then we may choose the number  $N$  of samples much smaller than  $D$  (but larger than  $M$ ), and we are still able to recover a polynomial  $f \in \Pi_\Gamma(M)$  exactly – at least with high probability.

Note that the theorem is non-uniform in the sense that it does not guarantee that a single sampling set  $\{x_1, \dots, x_N\}$  is good for all support sets  $T$  of a certain cardinality  $M$ . This slight drawback can actually be removed at the cost of introducing additional log-factors by using the concept of the uniform uncertainty principle introduced in [9, 12]. For an  $N \times D$  matrix  $A$

and  $M \leq D$  the *restricted isometry constant*  $\delta_M$  is defined as the smallest number such that for all subsets  $T \subset \{1, \dots, D\}$  with cardinality  $|T| \leq M$

$$(1 - \delta_M)\|x\|_2^2 \leq \|A_T x\|_2^2 \leq (1 + \delta_M)\|x\|_2^2$$

for all coefficient vectors  $x$  supported on  $T$ . Here  $A_T$  denotes the submatrix of  $A$  consisting of the columns indexed by  $T$ . In [9] the following general recovery theorem for BP was proved.

**Theorem 2.2.** *Assume that  $A$  is a matrix satisfying*

$$\delta_M + \delta_{2M} + \delta_{3M} < 1. \quad (2.4)$$

*Then every vector  $x$  with at most  $M$  non-zero entries can be recovered from  $y = Ax$  by BP.*

The above statement is deterministic, but unfortunately, it is hard to check that a deterministic matrix  $A$  has small enough  $\delta_M$ 's. So the strategy is to prove that a random matrix, in particular, our matrix  $\mathcal{F}_X$ , satisfies condition (2.4). Indeed, this was first achieved by Candès and Tao in [12] for partial discrete Fourier matrices  $\mathcal{F}_X$ , and later improved by Rudelson and Vershynin in [38]. Recently, Rauhut [34] used Rudelson and Vershynin's method to come up with an analog result for  $\mathcal{F}_X$  with samples  $X$  drawn from the continuous distribution on  $[0, 2\pi]^d$ . More precisely, if

$$N/\log(N) \geq C_\delta M \log^2(M) \log(D) \log^2(\epsilon^{-1}) \quad (2.5)$$

then with probability at least  $1 - \epsilon$  the restricted isometry constant of the  $N \times D$  matrix  $N^{-1/2}\mathcal{F}_X$  satisfies  $\delta_M \leq \delta$ , both for the continuous and discrete probability model, see [34, Theorem 3.2], [38]. The above condition is particularly good for small  $M$ , but is always satisfied if  $N \geq CM \log^4(D) \log^2(\epsilon^{-1})$ . (It seems that the  $\log^2(\epsilon^{-1})$ -factor can be improved to  $\log(\epsilon^{-1})$ , see [39]). We note that Candès and Tao obtained the condition  $N \geq CM \log^5(D) \log(\epsilon^{-1})$  (substitute  $\epsilon = cD^{-\rho/\alpha}$  in [12, Definition 1.12] of the uniform uncertainty principle and use [12, Lemma 4.3]).

Combining Theorem 2.2 and condition (2.5) gives a uniform result for recovery of sparse trigonometric polynomials by BP, see also [12, 34, 38]. Thus, a single sampling set  $X$  may be good for all  $f \in \Pi_\Gamma(M)$ , in particular for all support sets  $T$ .

Note that in [9] an extension of Theorem 2.2 to noisy and non-sparse situations is provided.

### 2.3 Recovery results for greedy algorithms

Let us first consider recovery by thresholding. Our theorem reads as follows.

**Theorem 2.3.** *Let  $f \in \Pi_\Gamma(M)$  with Fourier coefficients supported on  $T$  with  $|T| = M$ . Define its dynamic range by*

$$R := \frac{\max_{k \in T} |c_k|}{\min_{k \in T} |c_k|}.$$

*Choose random sampling points  $X = (x_1, \dots, x_N)$  according to one of our two probability models. If for some  $\epsilon > 0$*

$$N \geq CM R^2 \log(4D/\epsilon) \quad (2.6)$$

*then with probability at least  $1 - \epsilon$  thresholding recovers the correct support  $T$  and, hence, also the coefficients  $c$ . The constant  $C$  is no larger than 17.89.*

Hence, also with thresholding we can recover a sparse trigonometric polynomials from few random samples. Again the number of samples  $N$  is linear in the sparsity. However, compared to Theorem 2.1 for Basis Pursuit there is an additional dependence on the dynamic range  $R$ . Moreover, the above theorem is non-uniform in contrast to the available results for Basis Pursuit. This means that recovery is successful with high probability only for the given support set  $T$  and coefficient vector  $c$ . It does not guarantee that a single sampling set  $X$  is good for recovering all sparse trigonometric polynomials. Indeed, if  $N \leq CM^2$  it can be shown that given the random sampling set  $X$  there exists with high probability a coefficient vector  $c = c(X)$  supported on  $T$  for which thresholding fails [35] (compare also the next section).

Let us now consider Orthogonal Matching Pursuit. One can expect that this iterative strategy performs better than the simple Thresholding algorithm. In particular, the dependence of the number of samples  $N$  on the dynamic range  $R$  as in (2.6) should be removed. This is indeed indicated by the next theorem.

**Theorem 2.4.** *Let  $f \in \Pi_\Gamma(M)$  with coefficients supported on  $T$ . Choose random sampling points  $X = (x_1, \dots, x_N)$  according to one of our two probability models. If*

$$N \geq CM \log(8D/\epsilon)$$

*then with probability at least  $1 - \epsilon$  OMP selects an element of the true support  $T$  in the first iteration. The constant  $C$  is no larger than 32.62.*

Our numerical experiments indicate that this result should extend to all iterations so that finally the true sparse polynomial  $f$  is recovered. Unfortunately, we have not yet been able to analyze theoretically the further iterations of OMP because of subtle stochastic dependency issues.

Again the above theorem is non-uniform and we refer to the next section for a uniform result, which, however, requires much more samples. Similarly as for thresholding, one can show that if  $N \leq CM^{3/2}$  then with high probability there exists an  $M$ -sparse coefficient vector  $c$  depending on the sampling set such that OMP fails in the first step [35]. We remark, however, that if OMP selects a wrong element in some step it might nevertheless recover the right polynomial. Indeed, if after  $M' > M$  steps the true support  $T$  is contained in the recovered support  $T'$  then the coefficients in  $T' \setminus T$  will usually be set to 0, since with high probability  $\mathcal{F}_{T'X}$  is injective (see also Lemma 3.2 in [33]). A precise analysis when this situation occurs seems to be quite involved, however.

## 2.4 Uniform results based on coherence

Many results for OMP rely on the so called coherence parameter  $\mu$  [41, 42]. It measures the correlation of different columns of the measurement matrix  $\mathcal{F}_X$ . It requires that they have unit norm, so we define  $\tilde{\phi}_k := N^{-1/2}\phi_k$  resulting in  $\|\tilde{\phi}_k\|_2 = 1$ . Then  $\mu$  is defined as

$$\mu := \max_{j \neq k} |\langle \tilde{\phi}_j, \tilde{\phi}_k \rangle| = N^{-1} \max_{j \neq k} |\langle \phi_j, \phi_k \rangle|. \quad (2.7)$$

Reformulating Corollary 3.6 in [41] for our context yields the following result.

**Theorem 2.5.** *Assume  $(2M - 1)\mu < 1$ . Then OMP (and also BP) recovers every  $f \in \Pi_\Gamma(M)$ .*

**Proof:** The matrix  $\mathcal{F}_{TX}$  is injective for all  $T$  with  $|T| \leq M$  due to the condition on  $\mu$ , see e.g. [42, Proposition 4.3]. This ensures uniqueness of a polynomial  $f \in \Pi_\Gamma(M)$  having the vector  $(f(x_1), \dots, f(x_N))$  of sample values. The rest of the proof is the same as the one of Corollary 3.6 in [41]. ■

A similar result holds also for thresholding, see e.g. [22]. Again there is an additional dependence on the ratio of the largest to the smallest coefficients.

**Theorem 2.6.** *Assume  $(2M - 1)\mu < R^{-1}$  for some  $R \geq 1$ . Then thresholding recovers every  $f \in \Pi_\Gamma(M)$  whose Fourier coefficients  $c$  satisfy*

$$\frac{\max_{k \in \text{supp } c} |c_k|}{\min_{k \in \text{supp } c} |c_k|} \leq R.$$

So clearly, we need to investigate the coherence of the random matrix  $\mathcal{F}_X$ .

**Theorem 2.7.** *Let  $X = (x_1, \dots, x_N)$  be chosen according to one of our two probability models. Suppose that*

$$N \geq C(2M - 1)^2 \log(4D'/\epsilon), \quad (2.8)$$

*where  $D' := \#\{j - k : j, k \in \Gamma, j \neq k\} \leq D^2$ . Then with probability at least  $1 - \epsilon$  the coherence of  $\mathcal{F}_X$  satisfies*

$$(2M - 1)\mu < 1,$$

*and consequently OMP recovers every  $M$ -sparse trigonometric polynomial. The constant satisfies  $C \leq 4 + \frac{4}{3\sqrt{2}} \approx 4.94$ . In case of the continuous probability model it can be improved to  $C = 4/3$ .*

Of course, by Theorem 2.6 a similar result applies also to thresholding. Note that always  $D \leq D' \leq D^2$ . If  $\Gamma = \mathbb{Z}_m^d$  and  $X$  is chosen at random from the grid  $\frac{2\pi}{m}\mathbb{Z}_m^d$  then actually  $D' = D - 1$  because  $j - k$  is computed in the “periodic sense”.

In contrast to the results of the previous section the above theorem is uniform in the sense that a single sampling set  $X$  is sufficient to ensure recovery of *all* sparse signals. However, it clearly has the drawback that now the number of samples  $N$  scales quadratically in the sparsity  $M$ . Apart from perhaps the constant  $C$  and the logarithmic scaling in the dimension  $D$  condition (2.8) actually does not seem to be improvable if one requires exact recovery of *all* sparse trigonometric polynomials from a single sampling set  $X$ , see the remark after Theorem 2.4. So in this regard there is a crucial difference between Basis Pursuit and greedy algorithms. For certain applications a non-uniform recovery result might be enough and then greedy algorithms work well (and usually much faster than BP), see also the numerical experiments in Section 4. But if one requires uniformity (and speed is not a concern) then Basis Pursuit seems to be the method of choice.

## 2.5 Related Work

**BP with other measurement ensembles.** Other choices of the measurement matrix  $A$  (instead of  $\mathcal{F}_X$ ) were also investigated by several authors, see e.g. [12, 14, 15, 16, 17, 37, 38]. For instance, it was shown (see e.g. [1, 12, 13, 37, 38]) that an  $N \times D$  random matrix  $A$  with



independent Gaussian distributed entries (with variance  $N^{-1}$ ) has restricted isometry constant  $\delta_M \leq \delta$  with probability at least  $1 - \epsilon$  provided

$$N \geq C_\delta M \log \left( \frac{D}{M\epsilon} \right). \quad (2.9)$$

Similar estimates are possible for Bernoulli matrices, i.e., random matrices with independent  $\pm 1$  entries. Condition (2.9) is slightly better than (2.5). So for certain applications of compressed sensing Gaussian / Bernoulli matrices  $A$  might be useful. However, such “completely random” matrices have the disadvantage of not being structured, hence in contrast to the Fourier case no fast algorithms are available for matrix vector multiplication. Moreover, the “samples”  $y = Ax$  lack a physical meaning. Of course, this does not matter in encoding / decoding problems. However, there are possible applications where the samples result as the output of a physical measurement or an image reconstruction problem [8] and can really be interpreted as samples of a trigonometric polynomial. Clearly, in such a case one must use the Fourier matrix  $\mathcal{F}_X$ .

**Orthogonal Matching Pursuit.** Concerning OMP (or greedy algorithms in general) in connection with compressed sensing much less investigations were done so far. Gilbert and Tropp proved the following result [18, Theorem 2] (set  $\epsilon = 2D^{-p}$  to arrive at the formulation below).

**Theorem 2.8.** *Let  $x$  be an  $M$ -sparse vector in  $\mathbb{R}^D$ . Let  $A$  be an  $N \times D$  random matrix with independent standard Gaussian entries. Assume that for some  $\epsilon > 0$*

$$N \geq 8M \log(2D/\epsilon).$$

*Then Orthogonal Matching Pursuit reconstructs  $x$  from the given data  $y = Ax$  with probability exceeding  $1 - \epsilon$ .*

The method in [18] actually applies to more general types of random matrices  $A$ , in particular, to Bernoulli matrices. However, it is heavily used that the columns of  $A$  are stochastically independent, so unfortunately their approach cannot be applied directly to the random Fourier matrices  $\mathcal{F}_X$ .

The above theorem is non-uniform, i.e., recovery is successful with high probability only for a given sparse vector  $x$ . It cannot be guaranteed that with high probability a single Gaussian measurement matrix  $A$  suffices for all  $M$ -sparse vectors  $x$ . As in the Fourier case Theorem 2.8 becomes actually false when requiring reconstruction for all coefficients supported on a given set  $T$ , see e.g. [16, Section 7].

**Ordinary Matching Pursuit** Ordinary Matching Pursuit is slightly simpler than OMP. The difference lies in steps 5 and 6 of Algorithm 1. Instead of performing the orthogonal projection over the whole space of previously chosen elements  $\phi_k$ , the new iterate and residual are computed as

$$\begin{aligned} d_s &= d_{s-1} + \langle r_{s-1}, \phi_{k_s} \rangle e_{k_s}, \\ r_s &= r_{s-1} - \langle r_{s-1}, \phi_{k_s} \rangle \phi_{k_s}, \end{aligned}$$

where  $e_k$  denotes the  $k$ -th canonical unit vector. This step is clearly faster than the orthogonal projection step. However, in contrast to OMP it may now happen that some elements  $k_s$  are selected more than once. So even if MP picks a correct element  $k \in T$  in every step it usually has not yet reconstructed the right coefficients after  $M = |T|$  steps. However, Gribonval

and Vandergheynst showed that exponential convergence can be guaranteed under the same condition as for OMP in Theorem 2.5, i.e.,  $(2M - 1)\mu < 1$ , see [23, Featured Theorems 1 and 2]. Hence, the uniform recovery result Theorem 2.7 has also an immediate application to ordinary Matching Pursuit. Moreover, Theorem 2.4 is applicable as well, since the first step of ordinary Matching Pursuit and of Orthogonal Matching Pursuit coincide.

**Sublinear Algorithms for Sparse Fourier Analysis.** Finally, we would like to point out that in the series of papers [19, 20, 46, 47] an approach from [27] has been considered. The algorithms are based on so called isolation and group testing from theoretical computer science. It has been proven that one can recover a sparse trigonometric polynomial  $f \in \Pi_\Gamma(M)$ ,  $\Gamma = \mathbb{Z}_m$ , from  $N$  randomly chosen samples on  $\frac{2\pi}{m}\mathbb{Z}_m$  if

$$N \geq CM^\alpha \log^\beta(D) \log^\gamma(1/\epsilon) \log^\delta(R), \quad \text{with some } \alpha, \beta, \gamma, \delta \in \mathbb{N}.$$

These results have been generalized to some extent for multivariate trigonometric polynomials in [20, 46, 47]. Moreover, [20] indicates that  $N \geq C'M \log^{\beta'}(D) \log^{\gamma'}(1/\epsilon) \log^{\delta'}(R)$  samples suffice if the random sampling set possesses additional structure.

The striking point in these algorithms is the fact that the computation time scales “sub-linear” as  $\log^\beta(D)$  in the dimension  $D$ . However, numerical experiments in [46, 47] show even for small sparsities  $M$  a rather large crossover point with the classical FFT which scales as  $D \log(D)$ . It seems that in practice these algorithms require much more samples than OMP and BP, so that the main concern of these methods is speed rather than a minimal number of samples.

### 3 Proofs

A main tool for our proofs is Bernstein’s inequality, see e.g. [44, Lemma 2.2.9] or [3].

**Theorem 3.1.** *Let  $Y_\ell$ ,  $\ell = 1, \dots, N$ , be a sequence of independent real-valued random variables with mean zero and variance  $\mathbb{E}[Y_\ell^2] \leq v$  for all  $\ell = 1, \dots, N$ . Assume that  $|Y_\ell| \leq B$  almost surely. Then*

$$\mathbb{P}\left(\left|\sum_{\ell=1}^N Y_\ell\right| \geq x\right) \leq 2 \exp\left(-\frac{1}{2} \frac{x^2}{Nv + Bx/3}\right).$$

Bernstein’s inequality allows to prove the following concentration inequality.

**Lemma 3.2.** *Assume that  $c$  is a vector supported on  $T$ . Further, assume that the sampling set  $X$  is chosen according to one of our two probability models. Then for  $j \notin T$  and  $x > 0$  it holds*

$$\mathbb{P}\left(|N^{-1} \langle \mathcal{F}_{TX} c, \phi_j \rangle| \geq x\right) \leq 4 \exp\left(-N \frac{x^2}{4\|c\|_2^2 + \frac{4}{3\sqrt{2}}\|c\|_1 x}\right).$$

**Proof:** Note that

$$\langle \mathcal{F}_{TX} c, \phi_j \rangle = \sum_{k \in T} \sum_{\ell=1}^N c_k e^{i(k-j) \cdot x_\ell} = \sum_{\ell=1}^N Y_\ell$$

where

$$Y_\ell = \sum_{k \in T} c_k e^{i(k-j) \cdot x_\ell}, \quad \ell = 1, \dots, N.$$

Since the  $x_\ell$  have the uniform distribution either on the grid  $\frac{2\pi}{m}\mathbb{Z}_m^d$  or on the cube  $[0, 2\pi]^d$ , and since  $j \notin T$  it is easy to see that  $\mathbb{E}Y_\ell = 0$ . Furthermore, the random variables  $Y_\ell$  are independent and bounded,

$$|Y_\ell| \leq \sum_{k \in T} |c_k| = \|c\|_1.$$

Their variance is given by

$$\mathbb{E}[|Y_\ell|^2] = \mathbb{E} \left[ \sum_{k, k' \in T} c_k \overline{c_{k'}} e^{i(k-j) \cdot x_\ell} e^{-i(k'-j) \cdot x_\ell} \right] = \sum_{k, k' \in T} c_k \overline{c_{k'}} \mathbb{E}[e^{i(k-k') \cdot x_\ell}] = \sum_{k \in T} |c_k|^2 = \|c\|_2^2.$$

Hereby, we used that  $\mathbb{E}[e^{i(k-k') \cdot x_\ell}] = \delta_{k, k'}$ . Thus, we obtain

$$\begin{aligned} \mathbb{P}(|N^{-1} \langle \mathcal{F}_{TX} c, \phi_\ell \rangle| \geq x) &= \mathbb{P} \left( N^{-1} \left| \sum_{\ell=1}^N Y_\ell \right| \geq x \right) \\ &\leq \mathbb{P} \left( N^{-1} \left| \sum_{\ell=1}^N \operatorname{Re}(Y_\ell) \right| \geq x/\sqrt{2} \right) + \mathbb{P} \left( N^{-1} \left| \sum_{\ell=1}^N \operatorname{Im}(Y_\ell) \right| \geq x/\sqrt{2} \right). \end{aligned}$$

Applying Bernstein's inequality to  $\operatorname{Re}(Y_\ell)$  and  $\operatorname{Im}(Y_\ell)$ ,  $\ell = 1, \dots, N$ , (using that  $|\operatorname{Re}(Y_\ell)| \leq |Y_\ell|$  and  $|\operatorname{Im}(Y_\ell)| \leq |Y_\ell|$ ) we obtain

$$\begin{aligned} \mathbb{P}(|N^{-1} \langle \mathcal{F}_{TX} c, \phi_\ell \rangle| \geq x) &\leq 4 \exp \left( -\frac{1}{2} \frac{N^2 x^2 / 2}{N \|c\|_2^2 + \frac{1}{3\sqrt{2}} \|c\|_1 N x} \right) \\ &= 4 \exp \left( -N \frac{x^2}{4 \|c\|_2^2 + \frac{4}{3\sqrt{2}} \|c\|_1 x} \right). \end{aligned}$$

This concludes the proof. ■

### 3.1 Proof of Theorem 2.3

Thresholding recovers the correct support if

$$\min_{j \in T} |N^{-1} \langle \phi_j, \mathcal{F}_{TX} c \rangle| > \max_{k \notin T} |N^{-1} \langle \phi_k, \mathcal{F}_{TX} c \rangle|.$$

Observe that if  $l \in T$  then the triangle inequality yields

$$\begin{aligned} |N^{-1} \langle \phi_l, \mathcal{F}_{TX} c \rangle| &= |c_l + N^{-1} \langle \phi_l, \mathcal{F}_{(T \setminus \{l\})X} c_{T \setminus \{l\}} \rangle| \\ &\geq \min_{j \in T} |c_j| - \max_{j \in T} |N^{-1} \langle \phi_j, \mathcal{F}_{(T \setminus \{j\})X} c_{T \setminus \{j\}} \rangle|, \end{aligned}$$

where  $c_{T \setminus \{l\}}$  denotes the vector  $c$  restricted to the indices in  $T \setminus \{l\}$ . Thus, thresholding is certainly successful if

$$\max_{j \in T} |N^{-1} \langle \phi_j, \mathcal{F}_{(T \setminus \{j\})X} c_{T \setminus \{j\}} \rangle| < \min_{j \in T} |c_j|/2 \quad \text{and} \quad \max_{k \notin T} |N^{-1} \langle \phi_k, \mathcal{F}_{TX} c \rangle| < \min_{j \in T} |c_j|/2.$$

We conclude that the probability that thresholding is *not* successful can be upper bounded by

$$\begin{aligned} & \mathbb{P} \left( \max_{j \in T} |N^{-1} \langle \phi_j, \mathcal{F}_{(T \setminus \{j\})X} c_{T \setminus \{j\}} \rangle| \geq \min_{j \in T} |c_j|/2 \right) + \mathbb{P} \left( \max_{k \notin T} |N^{-1} \langle \phi_k, \mathcal{F}_{TX} c \rangle| \geq \min_{j \in T} |c_j|/2 \right) \\ & \leq \sum_{j \in T} \mathbb{P} \left( |N^{-1} \langle \phi_j, \mathcal{F}_{(T \setminus \{j\})X} c_{T \setminus \{j\}} \rangle| \geq \min_{j \in T} |c_j|/2 \right) + \sum_{k \notin T} \mathbb{P} \left( |N^{-1} \langle \phi_k, \mathcal{F}_{TX} c \rangle| \geq \min_{j \in T} |c_j|/2 \right). \end{aligned}$$

Applying Lemma 3.2 we can bound the probability that thresholding fails by

$$\begin{aligned} & 4|T| \exp \left( -N \frac{\min |c_j|^2/4}{4 \max \|c_{T \setminus \{j\}}\|_2^2 + \frac{4}{3\sqrt{2}} \max \|c_{T \setminus \{j\}}\|_1 \min |c_j|/2} \right) \\ & + 4(D - |T|) \exp \left( -N \frac{\min |c_j|^2/4}{4 \|c\|_2^2 + \frac{4}{3\sqrt{2}} \|c\|_1 \min |c_j|/2} \right) \\ & \leq 4D \exp \left( -N \frac{\min |c_j|^2}{16 \|c\|_2^2 + \frac{8}{3\sqrt{2}} \|c\|_1 \min |c_j|} \right) \\ & \leq 4D \exp \left( -N \frac{\min |c_j|^2}{16M \max |c_j|^2 + \frac{8}{3\sqrt{2}} M \max |c_j| \min |c_j|} \right) \end{aligned} \tag{3.1}$$

$$\leq 4D \exp \left( -\frac{N \min |c_j|^2}{M \max |c_j|^2} \frac{1}{16 + \frac{8}{3\sqrt{2}}} \right), \tag{3.2}$$

where  $\min |c_j|$  and  $\max |c_j|$  are always taken over  $j \in T$ . We note that in (3.1) the following obvious estimates were used,

$$\|c\|_2^2 \leq M \max_{j \in T} |c_j|^2, \quad \text{and} \quad \|c\|_1 \leq M \max_{j \in T} |c_j|.$$

Requiring that the term in (3.2) is less than  $\epsilon$  and solving for  $N$  shows the claim.

### 3.2 Proof of Theorem 2.4

As basic ingredient of the proof we will use that a submatrix  $\mathcal{F}_{TX}$  is well-conditioned under mild conditions on the number of sampling points  $N$  and sparsity  $M$ . The corresponding result is taken from [24]. It is based on an analysis in [33] of the expectation of the Frobenius norm of high powers of  $NI - \mathcal{F}_{TX}^* \mathcal{F}_{TX}$ , which uses a combinatorial argument based on set partitions and is inspired by [8].

**Theorem 3.3.** *Let  $T \subset \mathbb{Z}^d$  be of size  $|T| = M$  and let  $x_1, \dots, x_N$  be chosen according to one of our two probability models. Choose  $\epsilon, \delta \in (0, 1)$  and assume*

$$\left\lfloor \frac{\delta^2 N}{3eM} \right\rfloor \geq \log(c(\delta)M/\epsilon),$$

where  $c(\delta) = (1 - e^{-1}\delta^2)^{-1} \leq (1 - e^{-1})^{-1} \approx 1.582$ . Then with probability at least  $1 - \epsilon$  the minimal and maximal eigenvalue of  $\mathcal{F}_{TX}^* \mathcal{F}_{TX}$  satisfy

$$1 - \delta \leq \lambda_{\min}(N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX}), \quad \text{and} \quad \lambda_{\max}(N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX}) \leq 1 + \delta. \tag{3.3}$$

Written in a more compact form, if

$$N \geq C\delta^{-2}M \log(M/\epsilon)$$

then (3.3) holds with probability at least  $1 - \epsilon$ . It seems that also techniques developed by Vershynin [45] can be used to provide a version of Theorem 3.3.

Now let us turn to the proof of Theorem 2.4. (Orthogonal) Matching Pursuit recovers an element of the support  $T$  in the first step if

$$\frac{\max_{j \notin T} |\langle \phi_j, \mathcal{F}_{TX} c \rangle|}{\max_{k \in T} |\langle \phi_k, \mathcal{F}_{TX} c \rangle|} = \frac{\max_{j \notin T} |N^{-1} \langle \phi_j, \mathcal{F}_{TX} c \rangle|}{\|N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX} c\|_\infty} < 1.$$

Assume for the moment that we are on the event that the minimal eigenvalue of  $N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX}$  is bounded from below by  $1 - \delta$ , cf. Theorem 3.3. Then

$$\|N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX} c\|_\infty \geq M^{-1/2} \|N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX} c\|_2 \geq M^{-1/2} (1 - \delta) \|c\|_2.$$

Hence, we need to bound the probability that

$$\max_{j \notin T} |N^{-1} \langle \phi_j, \mathcal{F}_{TX} c \rangle| < \frac{1 - \delta}{\sqrt{M}} \|c\|_2.$$

Using Lemma 3.2 we can estimate the probability of the complementary event by

$$\begin{aligned} & \mathbb{P} \left( \max_{j \notin T} |N^{-1} \langle \phi_j, \mathcal{F}_{TX} c \rangle| \geq \frac{1 - \delta}{\sqrt{M}} \|c\|_2 \right) \\ & \leq \sum_{j \notin T} \mathbb{P} \left( |N^{-1} \langle \phi_j, \mathcal{F}_{TX} c \rangle| \geq \frac{1 - \delta}{\sqrt{M}} \|c\|_2 \right) \\ & \leq 4(D - M) \exp \left( -\frac{N}{M} \frac{(1 - \delta)^2 \|c\|_2^2}{4\|c\|_2^2 + \frac{4}{3\sqrt{2}} \|c\|_1 \|c\|_2 (1 - \delta) / \sqrt{M}} \right) \\ & \leq 4(D - M) \exp \left( -\frac{N}{M} \frac{(1 - \delta)^2}{4 + \frac{4}{3\sqrt{2}} (1 - \delta)} \right). \end{aligned}$$

In the last step, we applied the Cauchy-Schwarz inequality  $\|c\|_1 \leq \sqrt{M} \|c\|_2$ .

Altogether, the probability that OMP fails in the first step can be bounded by

$$\mathbb{P}(\lambda_{\min}(N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX}) < 1 - \delta) + 4(D - M) \exp \left( -\frac{N}{M} \frac{(1 - \delta)^2}{4 + \frac{4}{3\sqrt{2}} (1 - \delta)} \right). \quad (3.4)$$

Now choose  $\delta = 1/2$ . Then by Theorem 3.3 the first term above can be bounded by  $\epsilon/2$  provided

$$\left\lfloor \frac{N}{12eM} \right\rfloor \geq \log \left( \frac{2}{1 - 1/(4e)} M/\epsilon \right). \quad (3.5)$$

Further, requiring that also the second term in (3.4) be less than  $\epsilon/2$  yields

$$N \geq \left( 16 + \frac{8}{3\sqrt{2}} \right) M \log(8(D - M)/\epsilon). \quad (3.6)$$

Since always  $M \leq D$  there exists a constant  $C \leq 32.62$  such that

$$N \geq CM \log(8D/\epsilon)$$

implies that both (3.5) and (3.6) are satisfied. This concludes the proof.

**Remark 3.4.** Clearly, we would like to analyze also the further steps of OMP. However, starting with the second step the coefficients  $c^{(s)}$  (as well as the current support set  $T_s$ ) of the current residual  $r_s$  depend on the chosen random sampling points. In this situation Lemma 3.2 cannot be applied directly anymore and the analysis becomes more difficult.

### 3.3 Proof of Theorem 2.7

Let  $k, j \in \Gamma$  with  $k \neq j$ . Observe that

$$N^{-1} \langle \phi_j, \phi_k \rangle = N^{-1} \sum_{\ell=1}^N e^{i(j-k) \cdot x_\ell} = N^{-1} \sum_{\ell=1}^N Y_\ell^{(j-k)}, \quad Y_\ell^{(j-k)} := e^{i(j-k) \cdot x_\ell}.$$

The random variables  $Y_\ell^{(j-k)}$  obey  $\mathbb{E}[Y_\ell^{(j-k)}] = 0$ ,  $|Y_\ell^{(j-k)}| = 1$ , and  $\mathbb{E}[|Y_\ell^{(j-k)}|^2] = 1$ . Hence, the union bound and Bernstein's inequality applied to  $\text{Re}(Y_\ell^{(j-k)})$  and  $\text{Im}(Y_\ell^{(j-k)})$ , see also the proof of Lemma 3.2, yields

$$\mathbb{P}(\max_{\substack{j,k \in \Gamma \\ j \neq k}} N^{-1} \left| \sum_{\ell=1}^N Y_\ell^{(j-k)} \right| \geq x) \leq \sum_{\substack{j,k \in \Gamma \\ j \neq k}} \mathbb{P}(N^{-1} \left| \sum_{\ell=1}^N Y_\ell^{(j-k)} \right| \geq x) \leq 4D' \exp\left(\frac{-Nx^2}{4 + \frac{4}{3\sqrt{2}}x}\right). \quad (3.7)$$

Setting  $x = (2M - 1)^{-1}$  yields

$$\mathbb{P}((2M - 1)\mu \geq 1) \leq 4D' \exp\left(-\frac{N}{(2M - 1)^2} \frac{1}{4 + \frac{4}{3\sqrt{2}}}\right).$$

Requiring that the latter term be less than  $\epsilon$  and solving for  $N$  shows the claim.

In case of the continuous probability model (1) it is possible to improve the constant. In [32] a sharp moment bound (Khinchine's inequality) for the sum of the variables  $Y_\ell^{(j-k)}$  is provided, i.e.,

$$\mathbb{E}[|\sum_{\ell=1}^N Y_\ell^{(j-k)}|^{2p}] \leq p! N^p. \quad (3.8)$$

Markov's inequality yields for  $\kappa \in (0, 1)$ , see also [43, Proposition 15],

$$\begin{aligned} \mathbb{P}(N^{-1} \left| \sum_{\ell=1}^N Y_\ell^{(j-k)} \right| \geq x) &\leq e^{-\kappa N x^2} \mathbb{E} \left[ \exp \left( \kappa N^{-1} \left| \sum_{\ell=1}^N Y_\ell^{(j-k)} \right|^2 \right) \right] \\ &= e^{-\kappa N x^2} \sum_{p=0}^{\infty} \kappa^p N^{-p} \frac{\mathbb{E}[|\sum_{\ell=1}^N Y_\ell^{(j-k)}|^{2p}]}{p!} \leq e^{-\kappa N x^2} \sum_{p=0}^{\infty} \kappa^p = \frac{e^{-\kappa N x^2}}{1 - \kappa}. \end{aligned}$$

Using this estimate instead of Bernstein's inequality, the above derivation (3.7) yields

$$\mathbb{P}((2M - 1)\mu \geq 1) \leq \frac{1}{1 - \kappa} D' \exp\left(-\frac{N}{(2M - 1)^2} \kappa\right).$$

Choosing for instance  $\kappa = 3/4$  results in the slightly improved condition, cf. (2.8),

$$N \geq \frac{4}{3} (2M - 1)^2 \log(4D'/\epsilon).$$

If the sampling points are chosen at random from the grid  $\frac{2\pi}{m}\mathbb{Z}_m^d$  then the moment bound (3.8) remains valid only for  $p = 1, \dots, m-1$ . One still expects that also in this case the constant  $C$  can be improved (possibly depending on  $m$ ), but the method above is not applicable directly any more.

## 4 Implementation and Numerical Experiments

### 4.1 Analysis of the time complexity

Orthogonal Matching Pursuit, cf. Algorithm 1, contains two costly computations. Step 4 multiplies the adjoint measurement matrix  $\mathcal{F}_X^*$  with the current residual vector  $r_s$ . When drawing the sampling set from the lattice  $\frac{2\pi}{m}\mathbb{Z}_m^d$  or from the cube  $[0, 2\pi]^d$ , we use a zero padded fast Fourier transform (FFT) or the nonequispaced FFT [31, 25], respectively. In both cases, the total costs of this step in one iteration is  $\mathcal{O}(D \log D)$ . Note furthermore, that if the maximum in step 4 occurs at several indices the algorithm chooses one of them.

Step 5 solves in each iteration a least squares problem

$$\|\mathcal{F}_{T_s X} d_s - f\|_2 \xrightarrow{d_s} \min.$$

A straightforward implementation yields costs  $\mathcal{O}(MN^2)$  per iteration. We accelerated this step by a QR factorization of  $\mathcal{F}_{T_s X}$  computed from the factorization of  $\mathcal{F}_{T_{s-1} X}$ , cf. [5, pp. 132], which reduces the costs to  $\mathcal{O}(N^2)$  per iteration. Alternatively, we use the iterative algorithm LSQR [30] which solves the least squares problem with only  $\mathcal{O}(MN)$  floating point operations since the matrices  $\mathcal{F}_{T_s X}$  have uniformly bounded condition numbers, cf. Theorem 3.3.

Clearly, if OMP succeeds, the algorithm takes  $M$  outer iterations. Two reasonable choices for stopping criteria are a maximum number of iterations (assuming an upper bound on the sparsity  $M$  is known) or a residual tolerance  $\varepsilon$  (or a combination of both). In any case the algorithm will do no more than  $N$  iterations. Assuming  $M$  outer iterations, a reasonable sparsity  $M = \mathcal{O}(\sqrt{D})$ , and  $N = \mathcal{O}(M \log D)$  sampling points as suggested by Theorem 2.4, Algorithm 1 (using LSQR) has a total cost of  $\mathcal{O}(D^{1.5} \log D)$  arithmetic operations.

Thresholding, cf. Algorithm 2, avoids the iterative procedure and thus takes  $\mathcal{O}(D \log D)$  operations if we assume again  $M = \mathcal{O}(\sqrt{D})$  and  $N = \mathcal{O}(M \log D)$ , see also Theorem 2.3. Note that the computation of the inner products  $\mathcal{F}_X^* r_s$  becomes the dominant task in both schemes if  $M = o(\sqrt{D})$ .

### 4.2 Matlab toolbox for Thresholding, OMP, and BP

For the reader's convenience, we provide an efficient and reliable implementation of the presented algorithms for the univariate case in Matlab. Following the common accepted concept of *reproducible research*, all numerical experiments are included in our publicly available toolbox [26]. The toolbox comes with a simple version of the nonequispaced FFT.

All examples testing the Basis Pursuit principle use the optimization tools of CVX [21], L1MAGIC [36], or MOSEK [29], respectively. If the vector of Fourier coefficients is assumed to be real valued, the  $\ell^1$ -minimization problem is reformulated as a linear program, whereas for complex valued coefficients the corresponding second order cone problem is set up. CVX and MOSEK handle both problems but the matrix  $\mathcal{F}_X$  has to be stored explicitly and no fast matrix vector multiplications by FFTs can be employed. In contrast, L1MAGIC includes the

use of function handles to avoid this memory bottleneck and reduces the computation time from  $\mathcal{O}(DN)$  to  $\mathcal{O}(D \log D)$  when multiplying with the matrix  $\mathcal{F}_X$ . Unfortunately, the solver for equality constraint  $\ell^1$ -minimization of this package supports only real valued coefficients.

### 4.3 Testbed and examples

All numerical results were obtained on a Intel PentiumM with 1.6GHz, 512MByte RAM running OpenSUSE Linux kernel 2.6.13-15-default and MatLab7.1.0.183 (R14) Service Pack 3. Subsequently, we compare our implementation of Algorithm 1 and Algorithm 2 with different Basis Pursuit implementations for the univariate case.

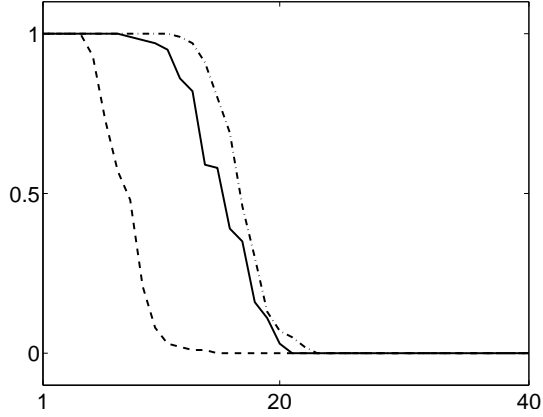
For given dimension  $D \in 2\mathbb{N}$  we set  $\Gamma = \{-D/2, -D/2+1, \dots, D/2-1\}$ . The support  $T \subset \Gamma$  of the Fourier coefficients is then chosen uniformly at random among all subsets of  $\Gamma$  of size  $M$ . We use two choices of (pseudo-)random Fourier coefficients. If not stated otherwise, the real as well as the imaginary part of the supported coefficients is drawn from a normal distribution with mean zero and standard deviation one. For some experiments with Thresholding, we choose coefficients with modulus one and draw the phase which from the uniform distribution on  $[0, 2\pi]$ .

The sampling points  $x_j$  are drawn uniformly from the interval  $[0, 2\pi]$  for the continuous probability model, denoted by NFFT subsequently. Within the discrete probability model a subset of size  $N$  is chosen uniformly among all subsets of  $\{0, \frac{2\pi}{D}, \dots, \frac{2\pi(D-1)}{D}\}$  with size  $N$ , denoted by FFT.

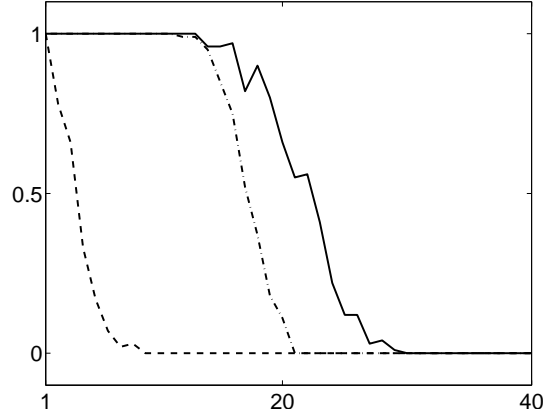
**Example 4.1.** *In our first example, we compare the ability of OMP, Thresholding, and BP to reconstruct sparse trigonometric polynomials with complex valued coefficients for the dimension  $D = 50$ . We draw a set  $T$  of size  $M \in \{1, 2, \dots, 40\}$  and  $M$  complex valued Fourier coefficients (with modulus one and a uniformly distributed phase in Figure 2(a)). Furthermore, we choose  $N = 40$  sampling points within the discrete or the continuous probability model. The samples  $(x_j, f(x_j))$ ,  $j = 1, \dots, N$ , of the corresponding trigonometric polynomial and the dimension  $D$  are the input for OMP, Thresholding, and BP. OMP and Thresholding use (updated) QR factorizations to solve the least squares problems, BP uses the MOSEK-package [29] to solve the second order cone problem. The output  $d_k \in \mathbb{C}$ ,  $k \in \Gamma$ , of all algorithms is compared to the original vector of Fourier coefficients. Repeating the experiment 100 times for each number  $M$  of non-zeros, we count how often each algorithm is able to reconstruct the given coefficients. Furthermore, the average CPU-time used by each algorithm with respect to the number of non-zero coefficients is shown. The same experiment is done for the dimension  $D = 1000$ .*

*As readily can be seen from Figure 2(d) and 2(f), Thresholding, OMP, and BP differ by orders of magnitude in their computation times. Note furthermore, that the number of outer iterations and hence the computation time grows linearly with the number of non-zeros for OMP. Regarding the rate of successful reconstructions, we observe the following: Thresholding fails already for a moderate number of non-zero coefficients, even if we use Fourier coefficients with modulus one, i.e., a dynamic range  $R = 1$ , cf. Figure 2(a). Basis Pursuit and Orthogonal Matching Pursuit achieve much better reconstruction rates, where surprisingly OMP outperforms BP for larger dimensions  $D$ , cf. Figure 2(e). Moreover, BP shows the same behavior for different dynamic ranges, the discrete and the continuous probability model, cf. Figure 2(a,b,c). In contrast, OMP seems to take advantage of a larger dynamic range of the coefficients in the FFT-situation, cf. Figure 2(b).*

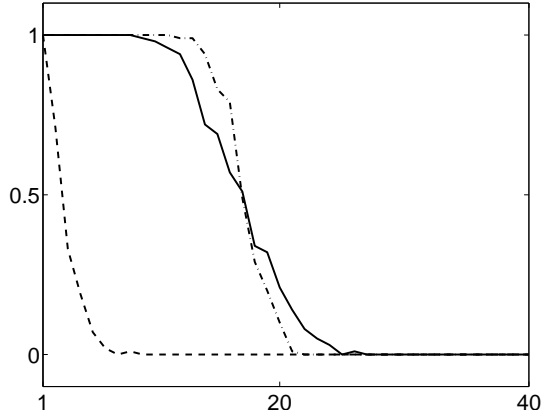




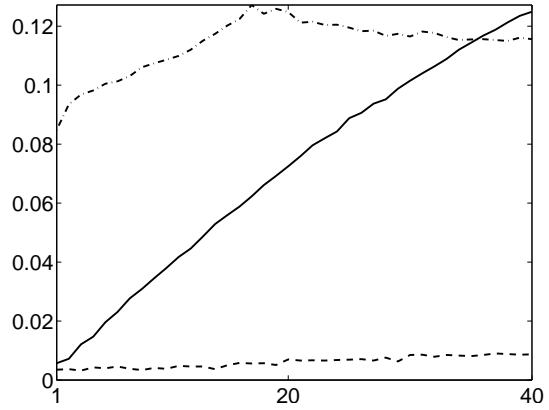
(a) Success rate vs. sparsity  $M = 1, \dots, 40$  for  $D = 100$ ,  $N = 40$ ,  $|c_k| = 1$  for  $k \in T$ , FFT.



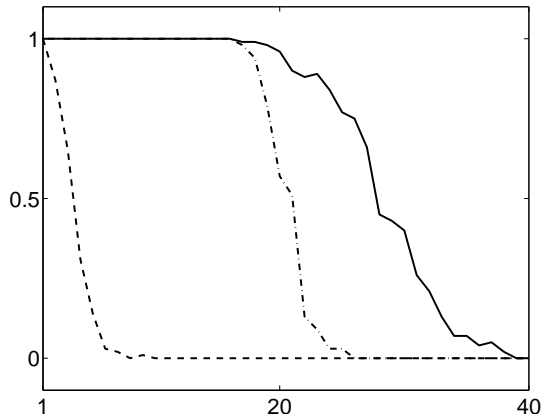
(b) Success rate vs. sparsity  $M = 1, \dots, 40$  for  $D = 100$ ,  $N = 40$ , FFT.



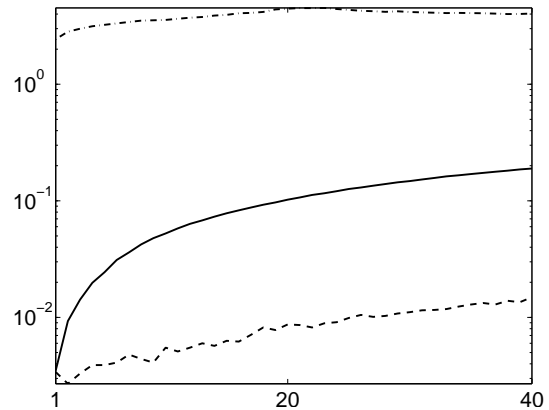
(c) Success rate vs. sparsity  $M = 1, \dots, 40$  for  $D = 100$ ,  $N = 40$ , NFFT.



(d) Computation time in seconds vs. sparsity  $M = 1, \dots, 40$  for  $D = 100$ ,  $N = 40$ , NFFT.



(e) Success rate vs. sparsity  $M = 1, \dots, 40$  for  $D = 1000$ ,  $N = 80$ , FFT.



(f) Computation time in seconds vs. sparsity  $M = 1, \dots, 40$  for  $D = 1000$ ,  $N = 80$ , FFT.

Figure 2: Success rate and computation time of Thresholding (dashed), OMP (solid), and BP (dash-dot) with respect to an increasing number  $M$  of non-zero Fourier coefficients. The number of samples  $N$  and the dimension  $D$  remain fixed; 100 runs have been conducted for each setting.

**Example 4.2.** *The purpose of the second experiment is the comparison for real-valued and complex-valued coefficients. We consider the Basis Pursuit principle exclusively. As in the previous example, we let  $D = 100$  but use only  $N = 30$  sampling points. Again, we repeat the experiment 100 times for each level of sparsity. The results in Figure 3 reveal the following: Besides the easier implementation and a speed up by 25 percent, the additional assumption to have real valued coefficients indeed saves roughly half of the needed samples to recover a sparse trigonometric polynomial.*

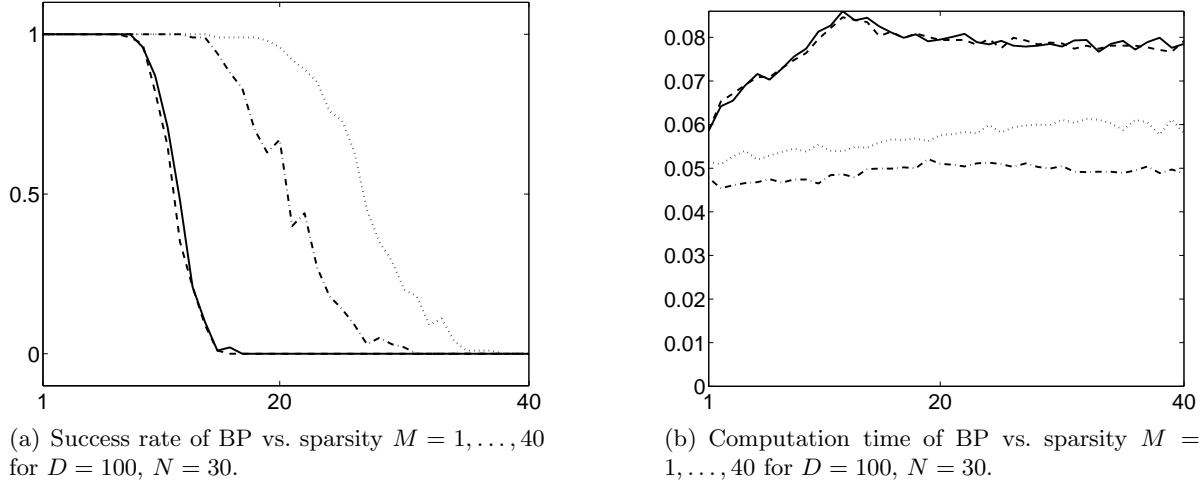
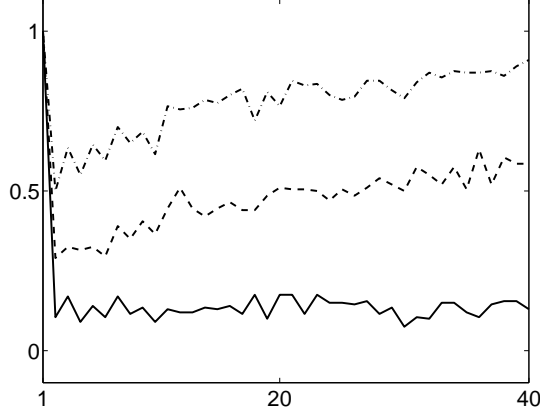


Figure 3: Success rate and computation time of Basis Pursuit for complex-valued coefficients with FFT (solid) and NFFT (dash-dot) and real-valued coefficients with FFT (dashed) and NFFT (dotted), with respect to an increasing number  $M$  of non-zero coefficients. The number of samples  $N$  and the dimension  $D$  remain fixed; 100 runs have been conducted for each setting.

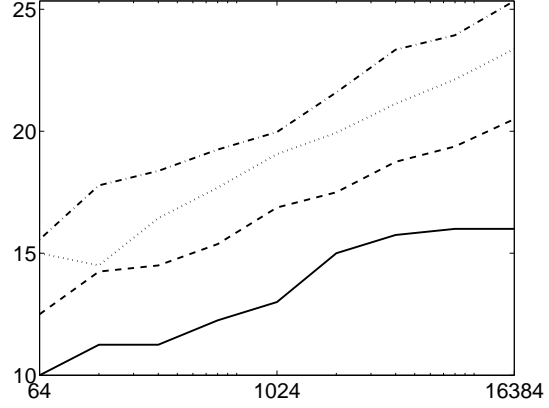
**Example 4.3.** *This example comments on the generalized oversampling factor  $\theta = N/M$  and its relation to the rate of successful reconstructions. For a fixed dimension  $D = 1024$  and a fixed generalized oversampling factor  $\theta > 0$ , we randomly draw a support set  $T \subset \Gamma$  of increasing sizes  $M = 1, \dots, 40$ , random Fourier coefficients (with modulus one and a uniformly distributed phase for Thresholding), and sampling sets in the continuous probability model of size  $N = \theta M$ . For 200 runs of each experiment, we count the number of perfect reconstructions by OMP after exactly  $M$  steps. As Figure 4(a) reveals, the success rate stays (almost) constant or might even increase slightly for an increasing number of non-zero coefficients if the ratio  $\theta = N/M$  remains constant, cf. Theorem 2.4 and 2.5.*

*In the second part of this example, we test Thresholding, OMP, and BP. Our concern is the dependence of the ratio  $\theta = N/M$  to reach a certain success rate when the dimension  $D$  varies. For an increasing dimension  $D = 2^6, 2^7, \dots, 2^{14}$ , we draw support sets  $T \in \Gamma$  of fixed sizes  $M = 4, 8, 16, 32$  and random Fourier coefficients (normal distribution for OMP and BP, modulus one for Thresholding). We then test for the smallest number  $N$  of continuously drawn samples, such that at least 90 percent (180 out of 200) of the runs result in a perfect recovery of the given Fourier coefficients. Figure 4(b-d) confirm the relation  $\theta = C_\epsilon \log_2(D)$  to reach a fixed success rate. In contrast to Thresholding, the constant  $C_\epsilon$  even decreases mildly for OMP and BP when using a larger number  $M$  of non-zero coefficients.*

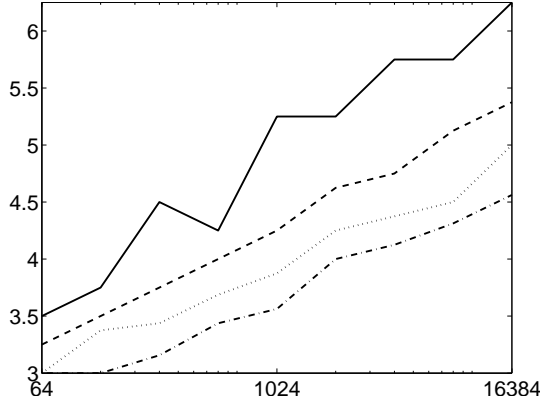
Both might indicate that the number of necessary samples is  $N = CM \log(D/(M\epsilon))$  rather than  $N = CM \log(D/\epsilon)$  for OMP and BP in an average case situation. Note furthermore, that the Gaussian ensemble indeed allows for reconstruction if this condition (2.9) is fulfilled.



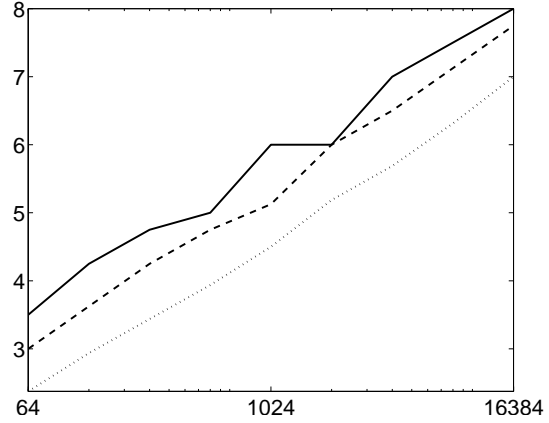
(a) Success rate of OMP vs. sparsity  $M = 1, \dots, 40$  for  $D = 1024$  and  $N = \theta M$  with  $\theta = 2.5$  (solid),  $\theta = 3$  (dashed), and  $\theta = 3.5$  (dash-dot).



(b) Oversampling factor  $\theta$  vs. dimension  $D$  for Thresholding ( $|c_k| = 1$  for  $k \in T$ ), 90% success.



(c) Oversampling factor  $\theta$  vs. dimension  $D$  for Orthogonal Matching Pursuit, 90% success.



(d) Oversampling factor  $\theta$  vs. dimension  $D$  for Basis Pursuit, 90% success.

Figure 4: Relation between the generalized oversampling factor  $\theta = N/M$  and the success rate for the continuous probability model. Figures (b)-(d): Oversampling factors  $\theta$  necessary to reach a success rate of 90 percent with respect to the dimension  $D$  and fixed numbers  $M = 4$  (solid),  $M = 8$  (dashed),  $M = 16$  (dotted), and  $M = 32$  (dash-dot) of non-zero coefficients.

**Example 4.4.** This example considers the computation time needed by OMP and BP for an increasing dimension  $D$ , a dependent number of non-zero coefficients  $M = \mathcal{O}(\sqrt{D})$ , and a number of samples  $N = \mathcal{O}(M \log D)$ . Constants are adjusted such that the used algorithms succeed in most cases in the reconstruction task; all methods, except L1MAGIC, are tested with complex coefficients. For small dimensions  $D$ , we draw the samples continuously and test OMP with the updated QR factorization and BP algorithms based on CVX, MOSEK, and L1MAGIC.

Furthermore, we test with a somewhat larger dimension and discrete drawn samples the algorithms: OMP with LSQR and explicitly stored matrices  $\mathcal{F}_{T_s X}$ , OMP with LSQR and FFT-based multiplications with  $\mathcal{F}_{T_s X}$ , BP using L1MAGIC and an explicitly stored matrix  $\mathcal{F}_X$  for  $D \leq 2^{12}$ , and BP using L1MAGIC and FFT-based multiplications with  $\mathcal{F}_X$ . The FFT-based multiplications are denoted by implicit in Figure 5 and need no additional memory, whereas storing the matrices explicitly needs  $\mathcal{O}(D \log D)$  bytes for OMP and  $\mathcal{O}(D^{1.5} \log D)$  bytes for BP.

Both OMP algorithms show a  $\mathcal{O}(D^{1.5} \log D)$  time complexity, whereas the scheme with explicit storage of  $\mathcal{F}_{T_s X}$  is a constant multiple faster. The Basis Pursuit algorithms are considerably slower in all cases. Moreover, the storage of the whole measurement matrix  $\mathcal{F}_X$  results in large memory requirements.

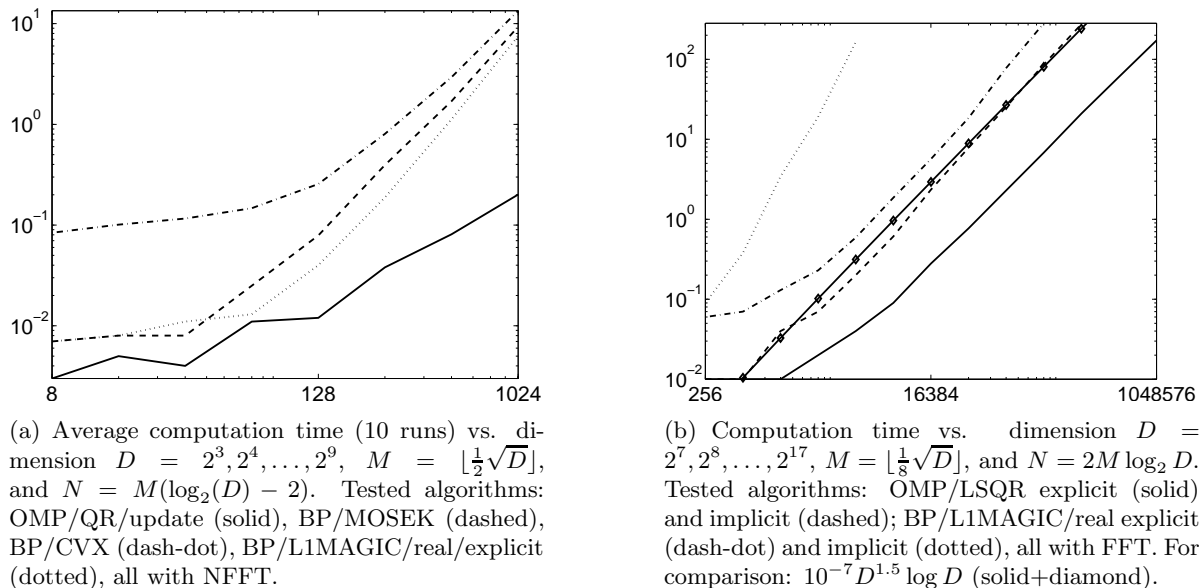


Figure 5: Computation time in seconds with respect to an increasing dimension  $D$  and dependent numbers  $M$  and  $N$  of non-zero coefficients and samples, respectively.

**Example 4.5.** Of course, stability under noise is important in practice. For Basis Pursuit this has been established already by Candès, Romberg and Tao in [9]. For Orthogonal Matching Pursuit we performed first numerical experiments in the following way. We corrupted the sample values with a significant amount of normal distributed noise. We observed that OMP usually finds the correct support set and makes only small errors on the coefficients, see Figure 6 and also [34]. Thus, it seems that also OMP is stable under noise – at least if the noise level is not too high. (Actually, we observed that for moderately higher noise as in our example in Figure 6, the reconstructed coefficient vector is significantly different from the original).

At last, note that both OMP and BP can be used to identify dominant frequencies from few samples – even if these frequencies are very high or if two (high) neighboring frequencies are present. Numerical tests showed for instance that the FFT based OMP can recover a signal consisting of 10 frequencies in  $\{-2^{18}, \dots, 2^{18} - 1\}$  – with  $k_1 = 2^{18} - 1 = 262143$ ,  $k_2 = 2^{18} - 2 = 262142$  being two of them – from 60 random samples.

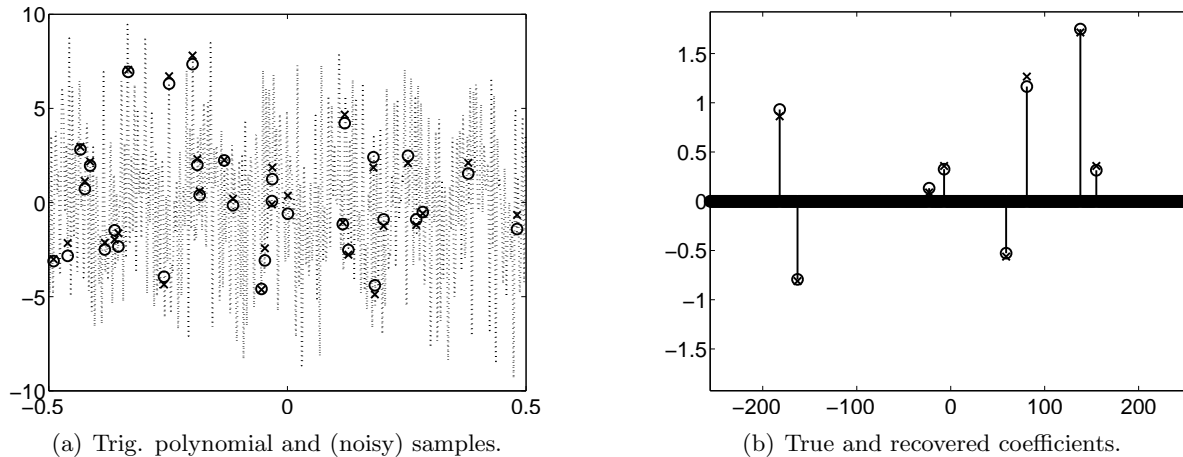


Figure 6: Trigonometric polynomial (real part) as in Figure 1 and 30 samples ( $\circ$ ). The samples are disturbed by Gaussian distributed noise with variance 0.2 ( $\times$ ) (resulting here in a PSNR of 15.6dB). Nevertheless OMP reconstructs the true support set of the coefficients, and the true coefficients ( $\circ$ ) themselves with small error ( $\times$ ).

## 5 Conclusions and future work

Our theoretical and numerical results indicate that both Basis Pursuit and greedy algorithms such as Orthogonal Matching Pursuit and Thresholding are well-suited for the problem of recovering sparse trigonometric polynomials from few random samples taken either on a grid (FFT) or on the cube (NFFT).

In practice, OMP obtains much better success rates than Thresholding. Moreover, our numerical results indicate that recovery success rates at reasonably small generalized oversampling rates  $N/M$  are very similar on generic signals for OMP and BP –although it is known that BP gives a uniform recovery guarantee while greedy algorithms only provide a non-uniform guarantee. As Theorem 2.4 just analyzes the first step of OMP it is still open to investigate theoretically the subsequent iterations. Due to stochastic dependency issues this does not seem to be straightforward, see Remark 3.4.

Both greedy methods are significantly faster than standard methods for Basis Pursuit. In contrast, the series of papers [19, 20, 46, 47] following [27] considers algorithms that allow for asymptotic running time proportional to  $\log^\beta(D)$  for some  $\beta$ . It would be very interesting to compare the two approaches numerically. Moreover, step 4 of Algorithm 1 might be replaced by so-called isolation and group testing techniques as used in [46, 47].

Typically, signals are not sparse in a strict sense but might still be well-approximated by sparse ones. Due to [9], recovery by Basis Pursuit is still possible with small errors provided that the restricted isometry constants are small. As was noted in Section 2.2, our random Fourier matrices  $\mathcal{F}_X$  satisfy this condition [12, 38, 34].

**Acknowledgments:** The first author is grateful for partial support of this work by the German Academic Exchange Service (DAAD) and the warm hospitality during his stay at the Numerical Harmonic Analysis Group, University of Vienna. The second author is supported by an Individual Marie Curie Fellowship funded by the European Union under contract MEIF-CT-2006-022811. He also acknowledges funding by the European Union’s Human Potential

Programme, under contract HPRN-CT-2002-00285 (HASSIP).

We would like to thank Emmanuel Candès, Justin Romberg, Ingrid Daubechies, Jared Tanner, Joel Tropp and Roman Vershynin for interesting discussions on the subject.

## References

- [1] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Preprint, 2007.
- [2] R. F. Bass, K. Gröchenig, Random sampling of multivariate trigonometric polynomials, SIAM J. Math. Anal. 36(3), 773 – 795, 2004.
- [3] G. Bennett. Probability inequalities for the sum of independent random variables, J. Am. Stat. Assoc. 57, 33–45, 1962.
- [4] A. Böttcher, D. Potts, Probability against condition number and sampling of multivariate trigonometric random polynomials, to appear in Electron. Trans. Numer. Anal., 2007.
- [5] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, 1996.
- [6] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [7] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20(1), 33 – 61, 1999.
- [8] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory 52(2), 489 – 509, 2006.
- [9] E. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math. 59(8), 1207 – 1223, 2006.
- [10] E. Candès and J. Romberg, Quantitative robust uncertainty principles and optimally sparse decompositions, Found. Computat. Math. 6(2), 227 – 254, 2006.
- [11] E. Candès and J. Romberg, Practical signal recovery from random projections, Preprint 2005.
- [12] E. Candès, T. Tao, Near optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inform. Theory 52, 5406–5425, 2006.
- [13] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51, 4203 – 4215, 2005.
- [14] D. L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52, 1289 – 1306, 2006.
- [15] D. L. Donoho, Y. Tsaig, Extensions of compressed sensing, Signal Proc. 86(3), 549 – 571, 2006
- [16] D. L. Donoho, For most large underdetermined systems of linear equations the minimal  $\ell^1$ -norm solution is also the sparsest solution, Comm. Pure Appl. Math. 59, 797 – 892, 2006.

- [17] D. L. Donoho, J. Tanner, Sparse nonnegative solutions of underdetermined linear equations, *Proc. Natl. Acad. Sci. USA* 102, 9446 – 9451, 2005.
- [18] A. Gilbert, J. Tropp, Signal Recovery from partial information via Orthogonal Matching Pursuit, revised Preprint, 2006.
- [19] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, M. Strauss, Near-optimal sparse Fourier representations via sampling, *Proc. STOC*, 2002.
- [20] A. Gilbert, S. Muthukrishnan, M. Strauss, Improved time bounds for near-optimal sparse Fourier representations, *Proc. SPIE*, 2005.
- [21] M. Grant, S. Boyd, Y. Ye, CVX: Matlab Software for Disciplined Convex Programming, Version 1.0RC3, <http://www.stanford.edu/~boyd/cvx>, 2007.
- [22] R. Gribonval, B. Mailhe, H. Rauhut, K. Schnass, and P. Vandergheynst, Average case analysis of multichannel thresholding, *Proc. ICASSP 2007*, 2007.
- [23] R. Gribonval and P. Vandergheynst, On the exponential convergence of matching pursuits in quasi-incoherent dictionaries, *IEEE Trans. Inform. Theory* 52, 255 – 261, 2006.
- [24] K. Gröchenig, H. Rauhut, B. Pötscher, Learning trigonometric polynomials from random samples and exponential inequalities for eigenvalues of random matrices, Preprint 2007.
- [25] J. Keiner, S. Kunis, D. Potts, NFFT Softwarepackage, C subroutine library, <http://www.tu-chemnitz.de/~potts/nfft>, 2002 – 2006.
- [26] S. Kunis, H. Rauhut, OMP-NFFT, MatLab-toolbox for orthogonal matching pursuit on sparse trig. polynomials, <http://www.tu-chemnitz.de/~skunis/software.php>, 2006.
- [27] Y. Mansour, Randomized interpolation and approximation of sparse polynomials, *SIAM J. Comput.* 24, 357 – 368, 1995.
- [28] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Sig. Proc.* 41, 3397 – 3415, 1993
- [29] mosek ApS, MOSEK optimization software, <http://www.mosek.com>.
- [30] C. Paige, M. A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. on Mathematical Software* 8, 43 – 71, 1982.
- [31] D. Potts, G. Steidl, M. Tasche, Fast Fourier transforms for nonequispaced data: A tutorial, In: *Modern Sampling Theory: Mathematics and Applications*, J. J. Benedetto and P. Ferreira (Eds.), Chapter 12, pages 249–274, 1998.
- [32] G. Peskir, Best constants in Kahane-Khintchine inequalities for complex Steinhaus functions, *Proc. Amer. Math. Soc.* 123, 3101–3111, 1993.
- [33] H. Rauhut, Random sampling of sparse trigonometric polynomials, *Appl. Comput. Harm. Anal.* 22, 16 – 42, 2007.
- [34] H. Rauhut, Stability results for random sampling of sparse trigonometric polynomials, Preprint, 2006.

- [35] H. Rauhut, R. Vershynin, On the impossibility of uniform recovery using greedy methods, in preparation.
- [36] J. Romberg, E. Candes, L1MAGIC, MatLab toolbox for basis pursuit, <http://www.acm.caltech.edu/l1magic>, 2006.
- [37] M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals, *Int. Math. Res. Notices* 64, 4019–4041, 2005.
- [38] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements, In: *Proc. CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.
- [39] M. Rudelson, R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Preprint, 2006.
- [40] J.F. Sturm, Using SeDuMi, a Matlab Toolbox for Optimization over Symmetric Cones, *Optim. Methods Softw.* 11-12, 625 – 653, 1999, now at <http://sedumi.mcmaster.ca/>.
- [41] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory* 50, 2231 – 2242, 2004.
- [42] J. Tropp, Topics in Sparse Approximation, PhD-Thesis, UT-Austin, 2004.
- [43] J. Tropp, Random subdictionaries of general dictionaries, Preprint, 2006.
- [44] A. Van der Vaart and J. Wellner, *Weak convergence and empirical processes*, Springer-Verlag, 1996.
- [45] R. Vershynin, Frame expansions with erasures: an approach through the non-commutative operator theory, *Appl. Comp. Harmonic Anal.* 18, 167–176, 2005.
- [46] J. Zou, A. Gilbert, M. Strauss, I. Daubechies, Theoretical and experimental analysis of a randomized algorithm for Sparse Fourier transform analysis, *J. Comp. Phys.* 211, 572 – 595, 2005.
- [47] J. Zou, A sublinear algorithm for the recovery of signals with sparse Fourier transform when many samples are missing, *Appl. Comput. Harm. Anal.* 22, 61 – 77, 2007.

Stefan Kunis  
 Fakultät für Mathematik  
 Technische Universität Chemnitz  
 Reichenhainer Str. 39  
 09107 Chemnitz, Germany  
[kunis@mathematik.tu-chemnitz.de](mailto:kunis@mathematik.tu-chemnitz.de)  
[www.tu-chemnitz.de/~skunis](http://www.tu-chemnitz.de/~skunis)

Holger Rauhut  
 NuHAG, Faculty of Mathematics  
 University of Vienna  
 Nordbergstr. 15  
 A-1090 Wien, Austria  
[holger.rauhut@univie.ac.at](mailto:holger.rauhut@univie.ac.at)  
[homepage.univie.ac.at/holger.rauhut](http://homepage.univie.ac.at/holger.rauhut)